

## **COSMO-RS: An Alternative to Simulation for Calculating Thermodynamic Properties of Liquid Mixtures.**

Andreas Klamt

COSMOlogic GmbH&Co.KG, Burscheider Str. 515, 51381 Leverkusen, Germany

e-mail: [klamt@cosmologic.de](mailto:klamt@cosmologic.de)

and

Institute of Physical and Theoretical Chemistry, University of Regensburg, Regensburg, Germany

Frank Eckert

COSMOlogic GmbH&Co.KG, Burscheider Str. 515, 51381 Leverkusen, Germany

e-mail: [eckert@cosmologic.de](mailto:eckert@cosmologic.de)

Wolfgang Arlt

Chair of Separation Science and Technology, University of Erlangen, Egerlandstr. 3, 91058 Erlangen, Germany,

e-mail: [wolfgang.arlt@cbi.uni-erlangen.de](mailto:wolfgang.arlt@cbi.uni-erlangen.de)

Corresponding author contact information:

Andreas Klamt

COSMOlogic GmbH&Co.KG, Burscheider Str. 515, 51381 Leverkusen, Germany

[klamt@cosmologic.de](mailto:klamt@cosmologic.de)

Tel.: +49-2171-731681, FAX: +49-2171-731689

## 1. Introduction

The liquid state is by far most important for chemistry and biology in general, and especially for chemical and biological engineering. For reaction and separation purposes it provides the advantage of permanent intensive molecular contacts with permanently changing partners, and molecules of different species often can be mixed and brought into contact in the liquid phase, if the solvents and mixtures are properly chosen. Therefore the liquid state is the preferred choice for chemistry chosen by biology and by chemical engineers. And hence the proper understanding of the interactions of molecules in liquid phases, and the proper knowledge of their liquid phase thermophysical properties from experiment or prediction, is an important prerequisite for any rational description and modification of chemical and biological processes.

While accurate experimental data is still an important and doubtlessly the most reliable source for thermophysical data of liquids, good experiments are expensive and time consuming, and thereby will ever be limited to a small subspace of possible combinations of solutes and pure or mixed solvents. Therefore theoretical or computational methods and models are important supplements for the exploration a larger space of new solutes and solvents. But unfortunately, just the same facts which make the liquid phase so preferable for biological and chemical engineering are making it extraordinary complicated for a theoretical description. The properties of molecules in liquid systems are influenced by all the interactions with the fluctuating neighbouring molecules, and their calculation requires an efficient sampling and thermodynamic averaging of all the possible arrangements of solute and solvent molecules. Strictly speaking, the properties of molecules in solution need to be calculated as combined thermodynamic and quantum mechanical expectation values of large ensembles of interacting

molecules. Obviously this is a hopelessly complex task, and therefore generations of researchers have tackled the problem with various degrees of simplification, empiricism and pragmatism.

Starting with more or less theoretically founded data interpolation schemes as

- $G^E$ -models like Wilson, NRTL, or UNIQUAC or equations of state, as widely used in process simulations (1),
- linear solvation free energy relationship models as the CLOGP method (2) which is massively used in bio- and medicinal chemical research,
- and ending with chemical engineering group-contribution methods as UNIFAC (3),

a large number of data driven, empirical models have been developed, which allow for property predictions if sufficient experimental data is available for the same or for structurally similar compounds. But such methods do not catch the molecular details of the underlying compounds and usually fail for new situations, especially for new classes of chemical compounds.

Physical chemists, and to a smaller extend also chemical engineers interested in a more fundamental description of molecular interactions of molecules therefore have developed atomistic simulation methods. These methods are based on force fields which are very detailed, but again highly parameterized schemes for the quantification of the different kinds of interactions of atoms. In combination with molecular dynamics (MD) or Monte-Carlo (MC) sampling techniques, nowadays this approach allows for the realistic simulation of liquid molecular ensembles, and for the calculation or thermodynamic averages of energies, volumes, and many other properties (4 - 7). The strength of such molecular simulation techniques is the ability to model the systems of interest with molecular detail. While especially the MC simulation techniques are rather efficient and mature, at least for pure

systems, the largest drawback of these methods consists in the limited accuracy and transferability of the force fields, resulting from the still too crude approximation of real molecules with their flexible electron distributions by ensembles of spherical atoms with fixed point charges. One way out would be the replacement of the force fields by the much more fundamental and hence much generally applicable quantum chemistry (QC) methods, which have become feasible meanwhile, but the system sizes and simulation times required for fluid phase simulations are still orders of magnitude too large for direct QC based thermodynamic simulations.

Currently, in QC solvents are usually simulated by continuum solvation methods (CSM). These are based on the oversimplified, but surprisingly successful approximation of the real solvent by a dielectric continuum of permittivity  $\epsilon$ . Nowadays this approach is implemented in most QC programs, exploiting the complete electrostatic information of the molecular electron density (for a review of CSMs see ref. 8). Klamt and Schüürmann developed a technical modification of the dielectric CSMs, the conductor-like screening model (COSMO), which replaces the dielectric boundary conditions by a much simpler scaled-conductor boundary condition (9). COSMO has become very popular due to the considerable reduction of the numerical demands and the increased numerical robustness. Nowadays a large part of the CSM calculations in QC are performed applying COSMO or small improvements of it, known as IEFPCM or SS(V)PE (10, 11). CSMs are parameterized on solvation energies of organic compounds, mostly for the solvent water. Usually each solvent has to be parameterized separately. CSMs neither provide any concept for mixtures, nor for the description of temperature effects, and thus they are not suited for chemical engineering applications, where temperature and composition variations are of crucial importance.

In 1995 Klamt introduced a novel combination of his COSMO model with statistical thermodynamics (12 - 14), called conductor-like screening model for realistic solvation (COSMO-RS). COSMO-RS provides a novel, rather direct and computationally very efficient pathway from quantum chemistry to fluid phase thermodynamics, enabling the usage of the predictive power and broad applicability of modern QC methods for the needs of chemical and biochemical engineers. COSMO-RS has been taken up enthusiastically by many chemical and bio-chemical engineers, resulting in a large number of publications reporting successful applications in many different areas, as well as in a few re-implementations of COSMO-RS, mostly by chemical engineering thermodynamics groups. In this review will describe the basic steps of the COSMO-RS theory and its application potential.

## **2. The conductor reference state**

From the perspective of dielectric theory, vacuum with  $\epsilon = 1$ , and a conductor, corresponding to a dielectric constant of  $\epsilon = \infty$ , are the ultimate extremes for the embedding of solutes.

While almost all computational chemistry methods start from the vacuum as reference, the COSMO-RS theory has introduced the state of molecules embedded in a conductor as conceptually fruitful starting point for fluid phase simulations.

The state of a molecule in a conductor can be very well calculated by QC calculations combined with COSMO. On the QC side it is recommended to use at least density functional (DFT) methods or MP2 ab initio theory, because simpler levels as Hartree-Fock or semi-empirical QC methods do not provide sufficiently accurate electrostatics. The concept of the QC/COSMO calculations required as input for COSMO-RS is rather simple:

- 1) Choose a start geometry of the solute X under consideration.

- 2) A cavity defining the boundary to the conductor and divided in sufficiently small segments is constructed around the solute.
- 3) An initial electron density is generated by the QC method.
- 4) The solute electrostatic potential arising from the atom nuclei and the electron density is calculated on the grid of cavity segment centres.
- 5) The conductor screening charge density  $\sigma$  is calculated from the conductor boundary condition that the total electrostatic potential arising from the solute and the polarization charges has to vanish on the entire surface.
- 6) The polarization charges are included as external charges into the next step of the QC iteration, resulting in a modified electron density. Hence steps 4 - 6 are repeated until self-consistency is achieved.
- 7) A correction for outlying charge errors (OCE) arising from the small portion of electron density outside the cavity is applied to the total energy and to the polarization charge densities.
- 8) The gradient of the total QC/COSMO energy is calculated analytically from the converged electron density and polarization charges, and a geometry step towards lower energy is performed. Steps 2 – 8 are iterated to self-consistency with respect to the solute geometry.

Finally this algorithm yields the self-consistent state, i.e. the energy, the electron density, the polarization charge densities, and the geometry of the molecule in a virtual conductor. We will call this “the COSMO state” further on. All relevant information about the COSMO state is stored in a COSMO file.

In addition to the choice of the QC method, the ambiguous steps in the COSMO workflow are step 2, the cavity construction, and the outlying charge correction in step 7. The cavity usually

is defined by the exterior of atom centred spheres. An important technical detail of the cavity construction is the method of smoothing along the crevices of intersecting spheres. While such smoothing is essential for the numerical stability and also for the physical consistency of the model, the development of a robust smoothing algorithm yielding a continuous cavity is tricky. The details of the smoothing algorithms differ considerably and this is a major source for differences in the results of different COSMO implementations. In addition, the OCE correction is treated differently in different implementations of COSMO. Although the COSMO algorithm is much less sensitive to the OCE than other dielectric CSMs, in the context of COSMO-RS it is strongly recommended to apply such OCE correction. More details on the cavity construction, smoothing algorithms and OCE correction can be found in the COSMO-RS book (13). In efficient implementations, QC/COSMO calculations usually only require about 20% more computation time than the corresponding gas phase calculations.

The conceptual value of the COSMO state as reference state for molecules in the liquid phase was only detected in the context of the COSMO-RS theory. In a perfect conductor all interactions are completely screened on the conductor interface, i.e. on the surface of the solute, by the conductor polarization charge density  $\sigma$ . Therefore, for an ensemble of molecules virtually swimming in a conductor, there are no intermolecular interactions. Hence each molecule in the COSMO state can just be considered individually. This decoupling makes the COSMO state a very attractive, clean reference state.

### **3. The molecular surface interaction concept of COSMO-RS**

In reality molecules are not swimming separately in a conductor bath, but do interact with each other. Starting from a reference ensemble of molecules swimming in a conductor, in which each molecule has its COSMO energy and its COSMO polarization charges, we can try

to approximate a closely packed liquid system by the iterative introduction of molecular contacts. Thus, from a thermodynamic point of view, the concept is a free enthalpy or Gibbs free energy concept of non-compressible fluids.

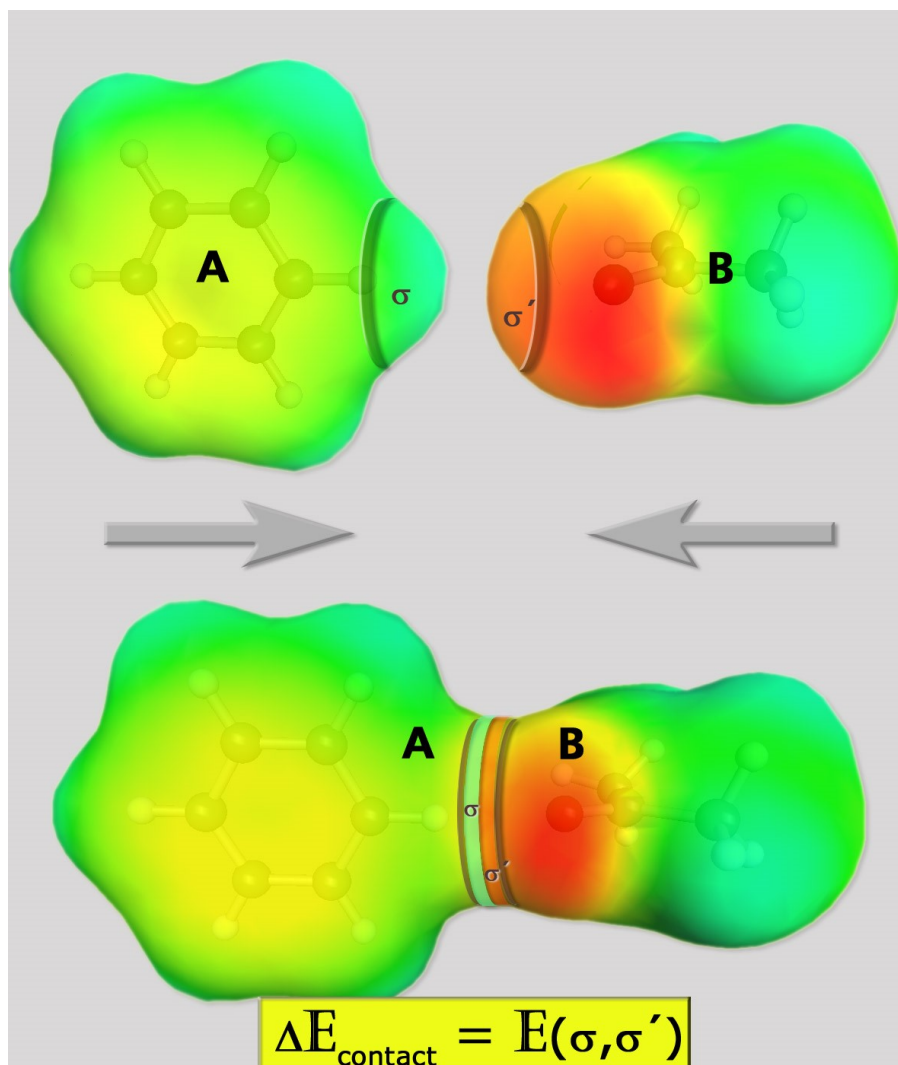


Figure 1: Schematic visualization of the COSMO-RS interaction concept

Let us first consider two molecules A and B of the ensemble, which may be different or of the same kind, and let us virtually reduce their distance until their COSMO surfaces touch each other, as schematically shown in figure 1. Since there are no interactions between the conductor-screened molecules, there is no change in energy during their approximation. Only in the final step, when the molecules get in direct contact and the conductor is removed between them on a certain contact area  $a_{\text{cont}}$ , there is a sudden energy change, because now the



AB complex is screened by the conductor as a whole. Assuming that the dispersive interactions of A and B experienced through the contact area  $a_{\text{cont}}$  are the same as they were in the conductor, the energy change due to the contact can be split into two contributions arising from the electrostatics and hydrogen bonding.

The electrostatic energy change is zero, if the conductor polarization charge densities  $\sigma$  and  $\sigma'$  of A and B on their contact surface areas are just opposite, i.e. if  $\sigma + \sigma'$  are zero, because then in total there is zero screening charge between the two molecules shortly before and after making the close contact. Hence the electrostatic energy and the polarization charge densities on the remaining conductor surface do not change at all during such electrostatically optimal contact. In reality, the molecules in a liquid ensemble will indeed have a tendency to make contacts of oppositely polar surface pieces. But due to thermal fluctuations or due to the lack of appropriate partners the sum of  $\sigma$  and  $\sigma'$  will not always be zero. Instead some residual charge density  $\sigma_{\text{misfit}}$  has to be taken into account. In general the electrostatic energy change of the AB complex compared to the individually screened molecules A and B going along with a contact of  $\sigma$  and  $\sigma'$  can be well approximated as

$$E_{\text{misfit}}(\sigma, \sigma') \cong a_{\text{cont}} e_{\text{misfit}}(\sigma, \sigma') = \frac{1}{2} a_{\text{cont}} \alpha_{\text{misfit}}' (\sigma + \sigma')^2 \quad (1)$$

This is just the energy required to neutralize the surface segment. The proportionality factor  $\alpha_{\text{misfit}}'$  depends on the surface  $a_{\text{cont}}$  and it takes into account the reduction of the neutralization energy arising from the electronic polarizability of the molecular environment, which usually is well represented by an optical permittivity of 2.

If the two surface segments which make the contact belong to a hydrogen bond donating and accepting atom, respectively, in a second step the donor hydrogen atom will get closer to the

acceptor and a hydrogen bond will be formed. In detail the exact quantification of the energy gain going along with the formation of such hydrogen bond requires a high level of quantum theory and it would not even be accurately described by the DFT methods usually employed for the COSMO calculations. Nevertheless, a bit more empirically, the hydrogen bond energy of a donor-acceptor contact gained after the contact of the COSMO surfaces can be reasonably well quantified based on the polarization charge densities  $\sigma$  and  $\sigma'$  of the contact segments, because hydrogen bonds are formed only between surface segments of sufficiently strong and opposite polarities  $\sigma$  and  $\sigma'$ . Since the hydrogen bond energy increases with the polarity of both, donor and acceptor, a simple expression as

$$E_{hb}(\sigma, \sigma') \cong a_{cont} e(\sigma, \sigma') = a_{cont} c_{hb}(T) \min(0, \sigma\sigma' - \sigma_{hb}^2) \quad (2)$$

already gives a reasonably accurate description of the hydrogen bond energy. The hydrogen bond threshold value  $\sigma_{hb}$  turns out to be in the range of 0.8-0.9 e nm<sup>-2</sup> (14). It should be noted that the coefficient  $c_{hb}$  depends on temperature. This describes the entropy loss going along with the formation of a hydrogen bond, resulting from the much stronger distance and orientation constraints of hydrogen bond contacts compared to the more floppy electrostatic contacts. Hence, strictly spoken,  $E_{hb}(\sigma, \sigma')$  should be considered as a hydrogen bond free energy. In summary we thus have expressed the contact interaction energy of the conductor embedded molecules A and B as a local interaction of the polarization charge densities  $\sigma$  and  $\sigma'$  of the two surface pieces of A and B making the contact:

$$E_{int}(A, B) \cong E_{int}(\sigma, \sigma') = a_{cont} e_{int}(\sigma, \sigma') \cong E_{misfit}(\sigma, \sigma') + E_{hb}(\sigma, \sigma') \quad (3)$$

With this expression for the first surface contact interaction energy we can virtually continue to build up one surface contact after the other, until the entire conductor is replaced by intermolecular contacts. If we neglect the Coulomb interaction energy of all the misfit charge densities with each other, which usually is small as long as the misfit charge densities are not correlated, we have thus constructed a liquid-like model of an ensemble of densely packed

molecules with a total interaction energy, which is expressed as a sum of local, pair-wise surface contact energies of COSMO polarization charge densities.

#### **4. The COSMO-RS statistical thermodynamics**

While the model considered above just represents a single liquid-like configuration of our molecular ensemble, the only way to calculate macroscopic thermodynamic properties of a liquid system is the application of statistical thermodynamics, i.e. to calculate thermodynamic averages over all possible configurations of the liquid. This is usually done employing MC or MD techniques in the context of force-field based energy expressions. It is obvious that such averaging always goes along with a loss of information. Motivated by the fact, that the COSMO-RS energy expression, in contrast to force-field energy expressions, does not explicitly depend on the full 3D geometry of the ensemble, but just is a sum of local surface contact energies, in COSMO-RS the complicated statistical thermodynamics of the nestled, 3D molecules is reduced to the much simpler statistical thermodynamics of just independently pair-wise interacting surface pieces.

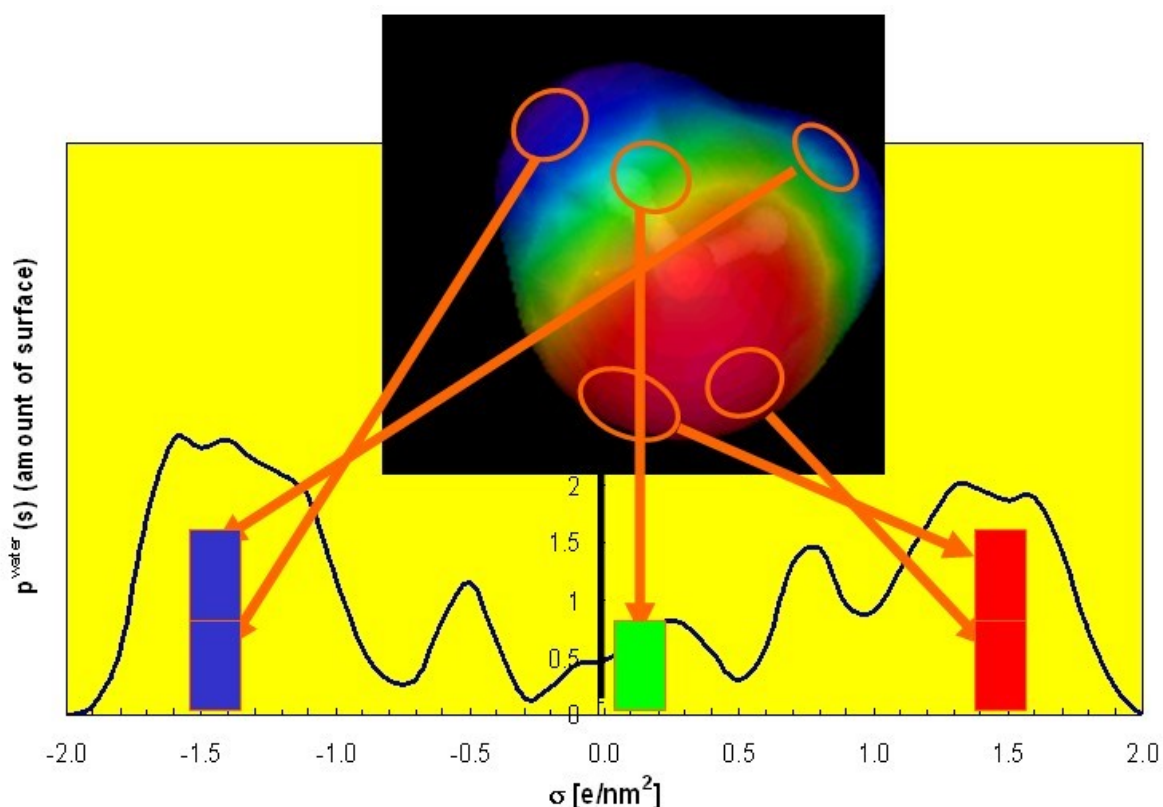
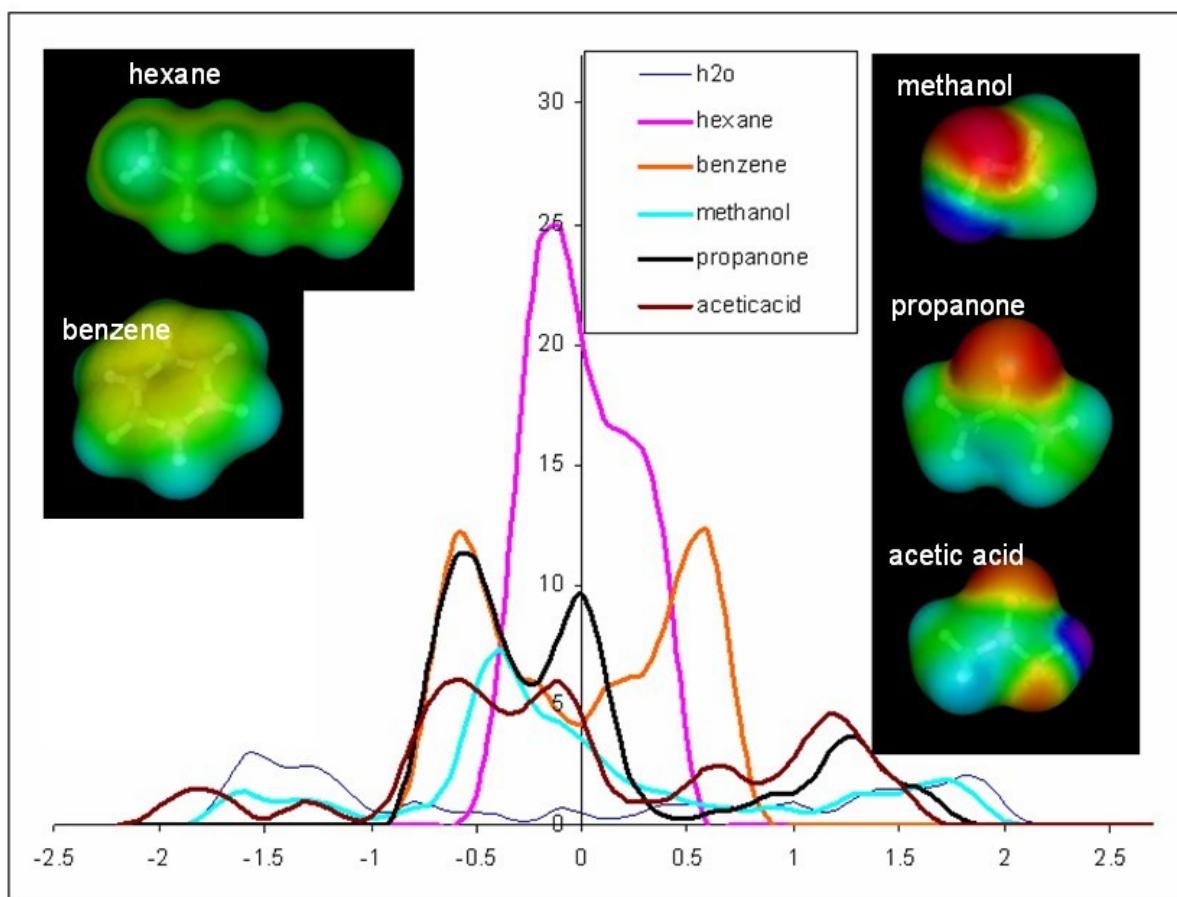


Figure 2:  $\sigma$ -profile of water (15)

As a preparation for this step, the concept of  $\sigma$ -profiles is introduced. The  $\sigma$ -profile  $p^X(\sigma)$  of a molecule  $X$  is the histograms of the molecular COSMO surface with respect to the polarization charge density  $\sigma$  as shown schematically for the water molecule in figure 2. In detail, it should be noted that a local average of  $\sigma$  is used for the generation of the polarization charge density, which is derived from the original polarization charge densities by averaging over the local neighbour segments using a Gaussian weight of width  $r_{av}$ . Although some re-implementations of COSMO-RS are using a slightly larger averaging radius, we consider  $r_{av} = 0.5 \text{ \AA}$  to be the best value. The  $\sigma$ -profile of water shows two pronounced peaks with maxima at about  $-1.5 \text{ e nm}^{-2}$  and  $+1.5 \text{ e nm}^{-2}$ , which result from the polar hydrogen atoms and the lone pair regions of the oxygen atom, respectively. In this picture we also introduce our colour coding of the COSMO cavities with respect to  $\sigma$ , in which deep blue stands for a surface area with strongly negative polarization charge density  $\sigma$ , i.e. for strongly positively polar parts of

the molecule, green for neutral parts of the surface and red for strongly positive parts of the COSMO surface, i.e. strongly negative molecular surface regions. The sign inversion between the polarization charge density  $\sigma$  and the molecular polarity just results from the fact that the conductor compensates the molecular polarity by opposite polarization charge density.



*Figure 3:  $\sigma$ -profiles of common compounds (15)*

In figure 3 we see the  $\sigma$ -surfaces and  $\sigma$ -profiles of a few other compounds. The  $\sigma$ -profiles turn out to be valuable finger prints of the molecular polarity. A detailed discussion of the various  $\sigma$ -profiles is given elsewhere (13).

Based on the individual  $\sigma$ -profiles of molecules we define the  $\sigma$ -profile of a solvent or a mixture as

$$p_s(\sigma) = \frac{\sum_i x_i p^i(\sigma)}{\sum_i x_i A^i} \quad (4)$$

where  $x_i$  is the mole fractions of component  $i$  and  $A^i$  is the respective COSMO surface area.

The surface normalization ensures that always the same amount of surface is considered in the statistical thermodynamics. From this solvent  $\sigma$ -profile we can calculate a solvent  $\sigma$ -potential by integration over all potential partners  $\sigma'$  in the solvent S:

$$\mu_s(\sigma) = -\frac{kT}{a_{eff}} \ln \int p_s(\sigma) \exp \left\{ -\frac{a_{eff}}{kT} (e_{int}(\sigma, \sigma') - \mu_s(\sigma')) \right\} d\sigma \quad (5)$$

Here the  $\sigma$ -potential  $\mu_s(\sigma)$  is the specific chemical potential, i.e. the chemical potential per surface area, of a piece of surface of polarity  $\sigma$  in the solvent ensemble characterized by the solvent  $\sigma$ -profile  $p_s(\sigma)$ . In simple words, the  $\sigma$ -potential is a characteristic function of a solvent or mixture S, specifying how much it likes surface area of polarity  $\sigma$ . Eq. 5 has been directly derived from the partition function of an ensemble of pair-wise interacting surface pieces of size  $a_{eff}$ . For a given interaction operator  $e_{int}(\sigma, \sigma')$  it represents the exact statistical thermodynamics of such ensemble. The effective contact area  $a_{eff}$  is a general parameter, which represents the size of a thermodynamically independent contact. When optimized within a COSMO-RS parameterization (12, 14), it takes the very reasonable value of 7.0 ( $\pm 0.5$ )  $\text{\AA}^2$ , corresponding to about 6 nearest-neighbour contacts for a molecule of the size of water.

While the term  $e_{int}(\sigma, \sigma')$  in eq. 5 represents the energetic costs of making a contact of  $\sigma$  with  $\sigma'$ , the appearance of  $\mu_s(\sigma')$  at this place represents the costs in free energy which is required in order to release a piece of surface of polarity  $\sigma'$  out of the other contacts and make it available for the contact with  $\sigma$ . Eq. 5 was independently derived for ensembles of pair-wise

interacting objects without any lattice concept. Later it has been shown to be equivalent to the exact solution of a quasi-chemical lattice approach. Indeed, Larsen and Rasmussen (16) once published a formal algorithm for the solution of the equation system of quasi-chemical lattice ensembles which is perfectly equivalent to eq. 5, but they did not introduce the conceptually very helpful interpretation of segment chemical potentials.

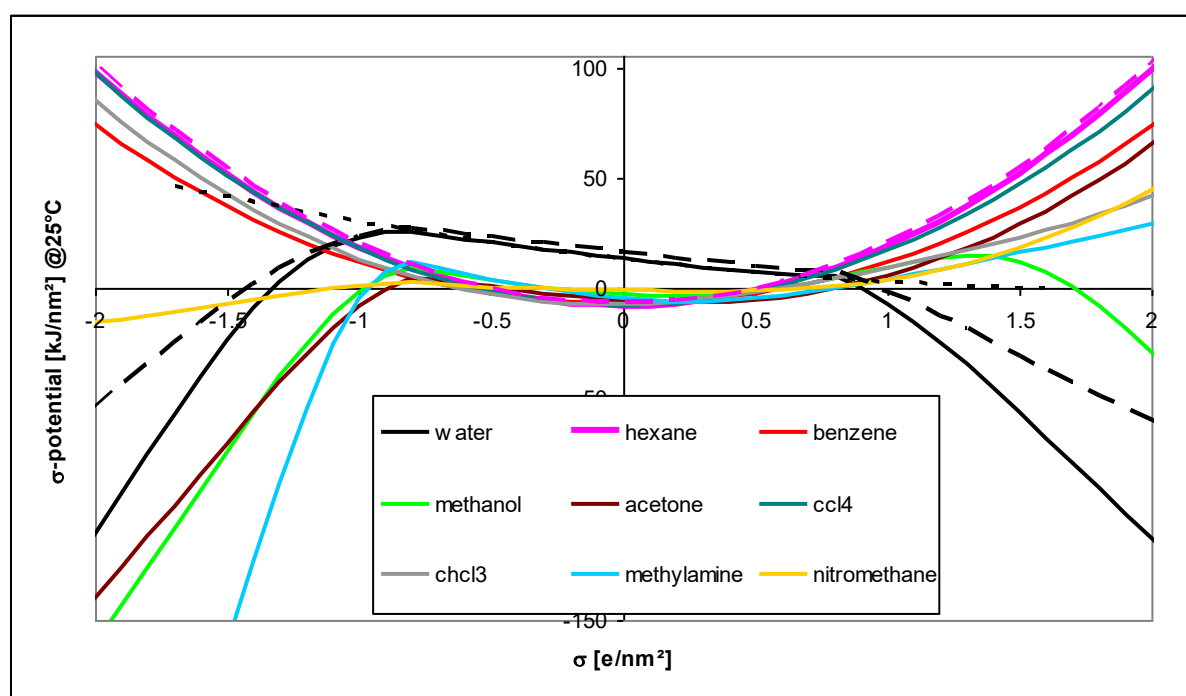


Figure 4:  $\sigma$ -potentials of representative solvents at 25°C (and 100°C, dashed curve)(15)

The  $\sigma$ -potentials of some representative solvents are shown in figure 4. The parabolic  $\sigma$ -potentials of hexane and benzene correspond to quasi-dielectric behaviour, but with a significant and essential difference between hexane and benzene, despite their almost identical macroscopic dielectric constant. This results from the broader  $\sigma$ -profile of benzene. The sharp decrease of some  $\sigma$ -profiles reflects the hydrogen bond affinity with respect to donors and acceptors, respectively. The lower  $\sigma$ -potential of methanol compared with water in the donor range ( $\sigma > 0.9 \text{ e/nm}^2$ ) reflects the larger donor affinity of methanol because it has

got only one hydrogen bond donor but two acceptor sites at the oxygen. As a consequence, the opposite can be observed for the acceptor affinity of methanol. Acetone shows a very strong hydrogen bond donor affinity, but no affinity for acceptors, because it does not have any donors which could form hydrogen bonds with acceptors. The  $\sigma$ -potential of non-polar surfaces in water is much higher than in most other solvents. This reflects the hydrophobic effect of non-polar molecular surface, resulting from the extremely small amount of non-polar surface area in water, and the strong interactions of all polar surfaces in water. It should be noted that COSMO-RS in this way does not only represent the hydrophobic effect in a natural way, but even correctly reproduces its entropic character, as has been shown in a study on the mutual solubilities of hydrocarbon compounds and water (17). Summarizing, we may conclude that the  $\sigma$ -profiles and  $\sigma$ -potentials are very specific solvent and solute characteristics, respectively, simultaneously describing the solvent and solute performance with respect to polar interactions, hydrogen bonding, and hydrophobic effects.

The final step of the COSMO-RS statistical thermodynamics is the calculation of the chemical potential of a molecule X solvated in solvent S by the integration of the  $\sigma$ -potential, i.e. by summation of the segment chemical potentials, over the surface of the solute X:

$$\mu_S^X = \int p^X(\sigma) \mu_S(\sigma) d\sigma + kT \ln \gamma_{comb}(X, S) \quad (6)$$

It must be noted that this is a quasi-chemical potential in the sense of Ben-Naim (18), i.e. that the trivial mole fraction term  $kT \ln(x)$  is omitted. In this article, we will further on use the expression "chemical potential" always in this sense. The first term in eq. 6 resulting from the interactions of solute X in solvent S and described via the COSMO-RS  $\sigma$ -potential, would be considered as the "residual part" of the chemical potential in the notation of chemical



engineering literature. The second term, the "combinatorial contribution", describes solute and solvent size dependence of the chemical potential, i.e. the fact that even in non-interacting liquids the chemical potential of a large molecule in one mole of small molecules is different from the chemical potential of a small molecule in one mole of large molecules, although for typical solute-solvent size ratios up to a factor 10 this difference is small. The COSMO-RS formalism itself would not describe such effects without a combinatorial term. Usually in COSMO-RS combinatorial terms rather similar to standard approaches from chemical engineering literature are used for this purpose, being based on the surface areas and volumes of solute and solvent molecules. In the context of COSMO-RS, these areas and volumes are routinely taken from the COSMO cavities, which are anyway available as a by product of the COSMO calculation. It should be noted, that the original size-correction term used in the COSMO-RS versions up to the year 2000 (12,14) was developed without the knowledge of the chemical engineering combinatorial terms and that it had a weak Gibbs-Duhem inconsistency, as detected first by G. Krooshof (private communication, 2000), and later pointed out by Lin and Sandler (19).

Eq. 6 represents the central equation of the COSMO-RS approach. With the restriction to incompressible liquids, it provides the chemical potential of an almost arbitrary solute X in an almost arbitrary liquid solvent or mixture as a function of temperature and concentration. Starting from the individual COSMO information of the pure compounds, it thus gives access to almost all thermodynamic liquid phase equilibrium properties of pure compounds and mixtures, e.g. activity coefficients, partition coefficients, enthalpies and entropies of mixtures and many more. From a chemical thermodynamics point of view, the COSMO-RS concept derived in the previous chapters is a free enthalpy or Gibbs enthalpy model of non-compressible fluids.

## 5. COSMO-RS gas phase and solid phase treatment

As explained before, COSMO-RS is a theory for molecules in the liquid phase. But many important thermodynamic properties as vapour pressures, partial pressures, Henry's law constants, etc., and VLE phase diagrams refer to the gas phase, and many others, especially solubilities of crystalline compounds, involve solid-liquid equilibria (SLE). For these cases COSMO-RS needs ways to estimate the involved free energy differences.

In contrast to simpler models, COSMO-RS can estimate the free energy of a compound in the ideal gas. The basic information for this is the quantum chemical energy difference between the vacuum reference state and the conductor reference state, i.e. the COSMO solvation energy at  $\epsilon = \infty$ . This includes all contributions to the free energy of phase transfer arising from the polarity and polarizability of the solute. We need to supplement this by a simple non-electrostatic contribution representing dispersive or vdW-interactions, a small correction for ring-structures, and a constant connecting the reference state in the ideal gas, which we chose as 1 bar, with the reference state in the liquid state, which we consider as 1 mol/mol. In summary, we yield the expression:

$$\mu_{Gas}^X = E_{vac}^X - E_{COSMO}^X + c_{disp}(T) \sum_{\alpha} A_{\alpha}^X \tau_{el(\alpha)} - \omega n_{ring}^X + \eta_0(T) \quad (7)$$

Nevertheless, if the pure compound vapour pressure is available for compound X, which often is the case, it is strongly recommended to use activity coefficients calculated by COSMO-RS in combination with the experimental vapour pressure for the evaluation of all VLE related properties, because this avoids the additional prediction error arising from eq. 7. It should be noted that eq. 7 does not take into account any gas phase non-idealities like dimerisation of organic acids. If such are considered as relevant, they need to be added externally starting from the ideal gas reference state.

The situation is similar, but somewhat worse, for SLE calculations. For the calculation of SLE the chemical potential difference of the compound between the liquid and the crystalline phase is required. In general, the prediction of the chemical potential of molecules in their crystalline phase is a rather unsolved task, because it requires the prediction of the crystal structure and free energy of the potential polymorphs as an initial step. The prediction of the free energy differences between the liquid and solid state from first principles is currently impossible. Therefore, for SLE calculations with COSMO-RS the usage of an experimental estimate for the fusion free energy  $\Delta G_{fus}^X$  based on the melting point and the heat of melting according to

$$\Delta G_{fus}^X(T) = -\Delta H_{fus}^X \left(1 - \frac{T}{T_{melt}^X}\right) + \Delta C p_{fus}^X (T_{melt}^X - T) - \Delta C p_{fus}^X T \ln \frac{T_{melt}^X}{T} \quad (8)$$

is strongly recommended. For high melting compounds as drugs and many fine chemicals, if possible the next order correction including the heat capacity change of fusion  $\Delta C p_{fus}^X$  should be used as well, but unfortunately this is rarely available from experiment. As a last resort, and due to the practical importance of drug solubility predictions, a heuristic estimate of  $\Delta G_{fus}^X$  based on COSMO-RS descriptors

$$\Delta G_{fus}^X(298K) \cong 12.2 V_{COSMO}^X - 0.76 N_{ringatom}^X + 0.54 * \mu_{water}^X \quad (9)$$

has been developed for usage in the context of solubility predictions of drug-like compounds (20). The three descriptors are the COSMO volume, the number of ring atoms, and the COSMO-RS chemical potential of the compound in water, representing size, rigidity, and a mixture of polarity and hydrogen bonding, respectively. This simple QSPR expression gives a reasonably accurate and robust description of  $\Delta G_{fus}^X$  at room-temperature for many neutral drug and pesticide data sets, but due to its purely heuristic nature this approximation cannot claim general applicability. It also should not be considered as part of the COSMO-RS method itself. Summarizing, for SLE calculations with COSMO-RS one only has the choice

between an often insufficient temperature extrapolation based on experimental melting information, or heuristic methods as the QSPR presented in eq. 9.

In combination with these estimates for the ideal gas and solid state chemical potentials, COSMO-RS can be used to calculate all kinds of VLE, LLE, and SLE properties.

## **6. COSMO-RS parameterizations and implementations**

COSMO-RS requires a number of adjusted parameters, although much less than other models used in chemical engineering. For example, the mod-UNIFAC method (21) would require roughly 10000 parameters, of which only about 50% are available. Furthermore, the COSMO-RS parameters usually are at most element specific and hence rather universally applicable. This may let COSMO-RS appear as almost *ab initio* from a chemical engineers perspective.

Indeed, not even the QC underlying COSMO-RS is parameter-free, because usually DFT is employed to generate the COSMO files containing the COSMO polarization charge densities. DFT calculations require the choice of a density functional, which always includes some empirical parameters. Since these are not fitted to fluid phase properties, they need not be considered parameters of COSMO-RS. Interestingly, the performance of almost all state of the art density functionals with respect to the quality of the COSMO-RS predictions is very similar. Semi-empirical QC methods as AM1, which was used in the first COSMO-RS publication, do not yield a sufficient electrostatic quality and hence meanwhile are considered as less suitable for COSMO-RS calculations. Conversely, MP2-COSMO calculations have been tested and do yield comparable quality with respect to COSMO-RS results as DFT/COSMO calculations, but they are not recommended due to the higher computational demands. A second degree of freedom in the COSMO calculations is the choice of the basis set. Different classes of basis sets are used or favoured in different QC programs. For this

kind of solvation calculations basis sets of triple-zeta polarization level are about the reasonable trade-off between accuracy and computation time. Larger basis sets usually do not yield better results anymore.

Of strong influence on the quality of COSMO-RS predictions is the set of radii used for the cavity definition in the QM/COSMO calculations. In the context of COSMO-RS, the COSMO radii usually are considered as element specific and as independent of the local bonding pattern and of the charge status of an atom. All COSMO-RS parameterizations and implementations are using the same set of COSMO radii published in the first quantitative COSMO-RS paper (14), where they have been obtained from a very compute intensive optimization. These element specific COSMO radii do show a rather systematic correlation with the Bondi radii (22) widely used in chemical engineering. The COSMO radius is quite systematically  $17(\pm 2)\%$  larger than the corresponding Bondi radius. This increase can be well understood, since the COSMO radii represent average distances to the neighbour cavity instead of nearest neighbour atom distances.

Other details of the cavity construction in the COSMO step are of less crucial importance. The "solvent radius", usually set equal to the hydrogen COSMO radius, is just a technical parameter for the smoothing of the COSMO surface in the intersection region of atomic spheres. The details of the smoothing algorithm can be of importance.  $\sigma$ -profiles generated from GEPOL cavities in the C-PCM algorithm (27, 28) show significantly different features than  $\sigma$ -profiles derived from the standard COSMO implementations, but often these differences are smoothed out during the thermodynamic averaging.

Some of the quite well defined parameters of COSMO-RS as the averaging radius  $r_{av}$ , the contact area  $a_{eff}$ , and the hydrogen bond threshold  $\sigma_{hb}$  have already been discussed in the

previous sections. The misfit energy coefficient  $\alpha'_{\text{misfit}}$  is of high importance for the COSMO-RS results. To first order it is quite reproducible and agrees reasonably with a crude estimate which can be derived from simple electrostatic arguments. But in detail it depends on the level of quantum chemistry used for the generation of the COSMO files, and it is able to compensate for small systematic under- or overestimation of molecular polarities going along with different quantum chemical methods and basis sets. The same is true for the details of the hydrogen bond coefficient  $c_{\text{hb}}(T)$ .

While the COSMO radii, the cavity construction details and the averaging radii usually are kept fixed since the first COSMO-RS parameterization (14), the other COSMO-RS parameters are often re-adjusted using a reasonably large set of various experimental data of great chemical diversity regarding solutes and solvents. For the COSMO*therm* line of parameterizations (25) currently a data set mainly consisting of room-temperature partition coefficients,  $\Delta G_{\text{hydration}}$  and vapour pressures is used, similar to, but meanwhile about twice as large as the data set used in the first quantitative COSMO-RS parameterization (14). The temperature dependent parameters are fitted only to vapour pressures, because for vapour pressure the largest amount of reliable temperature dependent data is available.

Apart from several more investigatory parameterizations on other QC levels, DFT functionals, and basis sets, the main three parameterizations supported by Klamt et al. are the BP-TZVP parameterization, the AM1-BP-SVP parameterization, and the DMOL<sup>3</sup> parameterization (25 - 28) parameterization. The first is the standard, suggested for all problems in which the number and size of the involved compounds allow for a full DFT geometry optimization. The required DFT/COSMO geometry optimizations can be most efficiently performed with TURBOMOLE (29), but a few other QC programs allow for the generation of equivalent BP-TZVP-COSMO files of almost identical quality. The AM1-BP-SVP parameterization has

been introduced for projects involving large numbers of larger, novel compounds, in which the computational demands for full DFT/COSMO geometry optimizations for all compounds may be prohibitive. Therefore the geometry optimization is performed on semi-empirical QC level, using the AM1 method (30) together with COSMO within the MOPAC7 program (31), supplemented by some molecular modelling corrections for a few known notorious geometry flaws of semi-empirical methods. The AM1/COSMO geometries combined with single point BP-SVP DFT/COSMO calculations do yield a rather good COSMO-RS parameterization, slightly less accurate than the BP-TZVP standard, but at only a few percent of the computation costs. DMol<sup>3</sup>-COSMO files have been used in the first quantitative COSMO-RS parameterizations. While being computationally somewhat less efficient than our current TURBOMOLE BP-TZVP standard, regarding accuracy the DMol<sup>3</sup>-COSMO files pertain to be an excellent basis for COSMO-RS parameterizations.

A number of re-implementations of COSMO-RS have been developed in the past years. The first was the re-implementation published by Lin and Sandler as COSMO-SAC (19), where SAC stands for "segment activity coefficients". While initially there was a debate on whether COSMO-SAC is a different model from COSMO-RS, it meanwhile is widely accepted that it is a straight re-implementation of the COSMO-RS concept with minor modifications. A second re-implementation was published by Grensemann and Gmehling as COSMO-RS(OL) (32). Banerjee et al. published another re-implementation of COSMO-RS and several chemical engineering applications (33), with special emphasis on ionic liquids. The most recent re-implementation of COSMO-RS was published by Pye et al. in the framework of the ADF program (34). While being based on different DFT-COSMO programs and methods, and having the one or other minor modification of the original COSMO-RS concept, all re-implementations confirm the overall robustness and reproducible quality of the COSMO-RS concept. Nevertheless, some comparisons of the COSMO<sup>therm</sup> parameterizations with

available re-implementations indicate that the latter do systematically underestimate large infinite dilution activity coefficients, most likely due to an under-representation of such cases in their parameterization data sets.

Within the line of COSMO-RS parameterizations available through the COSMO*therm* program (25), a number of additional features have been introduced on top of the basic COSMO-RS concept, the most important being the self-consistent treatment of multiple conformations (see below), the introduction of a second polarization charge density averaged over a larger vicinity of a segment in order to take into account some correlation effects of the polarization charge densities, and the introduction of non-additive vdW-interactions for some pairs of elements, especially for fluorine interactions. The latter are crucial for the ability to describe the non-ideal behaviour and miscibility gaps in alkane-perfluoroalkane mixtures. To our best knowledge, these improvements over the initial COSMO-RS concept are not available in the COSMO-RS re-implementations.

The overall accuracy of COSMO-RS, measured in the RMSD of transfer free energies or enthalpies of neutral compounds, has been  $\sim 1.7 \text{ kJ mol}^{-1}$  in the first quantitative DFT-based COSMO-RS parameterization, and it has improved to about  $\sim 1.3 \text{ kJ mol}^{-1}$  in recent COSMO*therm* parameterizations. The relative small improvements achieved in the past years, together with estimates of the electrostatic accuracy of DFT methods, indicate that this is about the limit of accuracy which can be expected for COSMO-RS based on DFT/COSMO. In detail the accuracy depends on the kinds of solutes and solvents, and on the properties considered. Obviously transfer free energies and equilibrium constants between polar and non-polar solvents, or polar solvents and gas-phase are subject to larger errors than activity coefficients in mixtures of similar liquids. Hydrogen bonding systems tend to show larger errors, due to the more empirical nature of the hydrogen bond interaction expression



compared to misfit energy expression. Nevertheless, in general COSMO-RS is quite good in predicting aqueous systems.

In a most recent validation the COSMOtherm implementation of COSMO-RS was shown yield an error of only  $2.0 \text{ kJ mol}^{-1}$  (MUE) on a dataset of almost 2500 free energies of solvation (35), which had been collected for the parameterization of the SM8 solvation model (36). Although not trained on this dataset, COSMOtherm still outperformed SM8 and other solvation models tested on this dataset. It should be noted that a number of data of questionable experimental quality were included in this dataset and contributed considerably to the average error.

## 7. Conformations

In the initial versions of COSMO-RS, and in the re-implementations published so far, each compound is just represented by a single COSMO file. While that is not a problem for rigid and for many small compounds, it becomes problematic for flexible molecules, which may have several or even a very large number of meta-stable geometries, so called conformations. Especially if the relative orientation of polar functional groups differs in the conformations, or if intramolecular hydrogen bonds can be formed, the energies and  $\sigma$ -profiles of such conformations can be significantly different. This may lead to the situation that the free energy of one conformation is lowest in one solvent, and another conformation is lowest in a second solvent. Therefore, for a consistent treatment it is necessary to take into account not just one conformation, but an ensemble of conformations, including all those which may be of low free energy in polar or non-polar solvents. Because the total free energy of each conformation in any solvent or mixture  $S$  is available within COSMO-RS as a sum of its DFT/COSMO energy in the conductor reference state and its individual pseudo-chemical potential according to eq. 6, the relative weight of each conformation follows from Boltzmann

statistics, and all thermodynamic properties of a multi-conformational compound can be evaluated as the corresponding averages (13). Since the chemical potentials depend on the average  $\sigma$ -profile of a system S, and since the latter via the Boltzmann weights depends on the chemical potentials, an additional self-consistency loop is required for the thermodynamically consistent treatment of systems with multi-conformational compounds. The only COSMO-RS implementation currently allowing for such consistent multi-conformational treatment of flexible compounds is the COSMO*therm* program.

For molecules with several rotatable bonds the search for the relevant conformations required for a consistent treatment can be very demanding. An automated procedure named COSMO*conf* has been recently developed by Klamt and co-workers (37). In contrast to standard conformational search tools it is especially designed for finding the low energy conformations in both, polar and non-polar solvents. Another strategy for the conformation generation and selection for COSMO-RS calculations, involving Monte-Carlo simulations, has been published by Arlt and co-workers (38).

## **8. Auxiliary tools and technical extensions of COSMO-RS**

For the efficient usage of COSMO-RS, databases of pre-calculated COSMO files for common solvents and other common compound classes are very valuable. The largest such database, including multi-conformational COSMO file for almost 5000 compounds, is COSMO*base* (39), and its extension to ionic liquids, COSMO*base*IL, which holds the COSMO files for most widely used anions and cations used for ionic liquids. Another database of about 2000  $\sigma$ -profiles has been developed in the context of COSMO-SAC (40) based on DMol/COSMO calculations. For most compounds, this just takes into account a single conformation per compound and is still not completely consistent regarding the choice of the conformations.

As any  $G^E$ -model, COSMO-RS itself is limited to incompressible, dense liquids. Several groups presented combinations of COSMO-RS with EoS or hole theory (40 – 47), in order to extend its applicability towards critical or near-critical liquids. Currently there is not sufficient experience in order to decide which of these combinations is most generally applicable.

As described above, the COSMO-RS method was mainly developed for the prediction of fluid phase equilibrium properties of molecules in bulk, homogeneous liquids. Recently it has been extended to the simulation of molecules at liquid-liquid and liquid vapour interfaces (13) by using the  $\sigma$ -potential  $\mu_S(\sigma)$  of phase S for that part of the solute surface which is in phase S, and the  $\sigma$ -potential  $\mu_{S'}(\sigma)$  of the other phase S' for that part of the solute surface which is in S'. By sampling all possible position, orientations and conformations this leads to the partition sum of the solute at the interface, and just provides useful information about the binding of compounds to interfaces. Generalization of this idea to multiple phases and consideration of micellar systems as layered liquids led to the development of the COSMOmic extension (48) of COSMO-RS, which allows for the prediction of free energies and the partition behaviour of compounds in micellar systems. It has been validated on the biologically important case of bio-membranes, i.e. aqueous DMPC bilayers, where it does not only well predict without any adjusted parameter the bio-membrane water partition coefficients, but also provides an efficient access to the free energy profiles of compounds for the passage through the membrane which provide important information for the cell permeability of drugs and other physiologically important compounds.

While COSMO-RS itself already is quite generally applicable to all partition problems involving chemically well defined phases, a slightly more empirical extension, the  $\sigma$ -moment approach, was required in order to apply it to many important partitioning and adsorption problems, in which one or both phases are less well defined. Although derived from theory,

the  $\sigma$ -moments have been shown to be strongly related to Abraham's empirical solvation parameter approach (49). The  $\sigma$ -moment approach has been successfully applied to such diverse problems as adsorption to activated carbon, blood-brain partitioning, intestinal absorption, soil-sorption, adsorption to cotton or hair, and many others ( 50 - 53).

While in its original form COSMO-RS can only be applied to molecules of limited size which can be treated by quantum chemistry, it has been extended to large periodic structures as thermoplastic polymers or crystal surfaces has been enabled by two different extension. One is the usage of the periodic boundary conditions during the DFT/COSMO calculations as they are enabled in the DMol program (54), and the other is the atom-weight technology implemented in the COSMOtherm code which allows for cutting the COSMO screening charges of the relevant repeat unit out of COSMO files of oligomers or surface clusters. Thus COSMO-RS can be applied to the prediction of solubilities in polymers, and to the calculation of free energies of crystal faces in solution (55).

In order to overcome the need for the potentially time-consuming DFT/COSMO calculations, the COSMO*frag* method (56) has been developed. Within less than a second this generates an approximate  $\sigma$ -profile of a new compound from a huge database of pre-calculated COSMO-files of diverse compounds, which can then be used for most kinds of COSMO-RS applications. COSMO*frag* can be very useful in large scale screening applications as they especially appear in drug design.

Due to the large information content and the general importance of  $\sigma$ -profiles for the physico-chemical and physiological behaviour of chemical compounds, it is often interesting to analyze or screen compounds under the aspect of  $\sigma$ -profile similarity, especially for drug design. For this purpose a robust  $\sigma$ -profile similarity measure and a fast screening technology

have been developed and published under the name *COSMOsim* (57), allowing for the efficient screening of millions of compounds. An extension of *COSMOsim* to 3D similarity is under development.

Although being almost straightforward application of COSMO-RS, it is worth mentioning the extension to the calculation of protonation and dissociation free energies and hence  $pK_a$  values in aqueous and non-aqueous solvents (58 - 60), which unfortunately requires some empirical scaling. Nevertheless, it can provide useful information about protonation equilibria of demanding and complicated compounds, which often are of large technical or physiological importance.

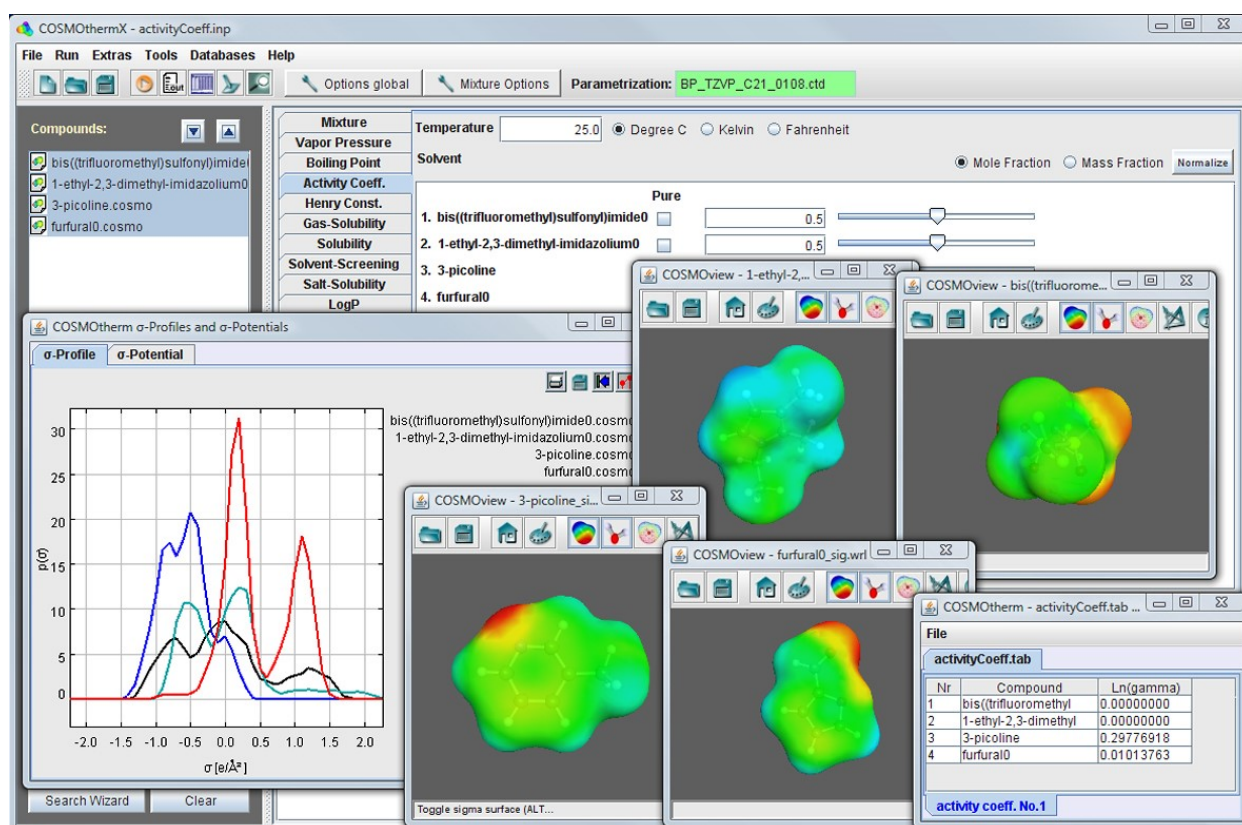


Figure 5: Screenshot of a typical COSMOtherm application showing  $\sigma$ -surfaces,  $\sigma$ -profiles, and activity coefficient output for 2 solutes in an ionic liquid.

Last not least the graphical user interface COSMO*therm*X (see figure 5) should be mentioned, which has become an important integration tool, making COSMO-RS calculations simple, graphic and easy to use even for non-experts, beginners and students.

## **9 Application range and user groups**

Since its original development of COSMO-RS, which was mainly focussed on solvation free energies, vapour pressures, and partition coefficient (12,14), COSMO-RS has been used for the prediction of a wide range of fluid properties by many different user groups. It is impossible to report all applications here. Therefore we will try to give a representative overview.

The most important application area has become chemical engineering thermodynamics, as it is required for process design and development. While being out of the focus of the original developers, this application area was opened by the pioneering application of COSMO-RS to the calculation of binary VLE diagrams by Clausen and Arlt in 1998 (61, 62), demonstrating the striking predictivity of COSMO-RS in such applications, although it was not developed nor parameterized for binary mixtures. Starting with BASF, Dupont, Bayer and Degussa (63 - 66), many chemical engineering groups in chemical companies shortly after recognized the predictive potential of COSMO-RS, especially for more demanding compounds, for which other methods fail, and they started to use it in chemical process design and development, especially for the screening and design of solvents and entrainers. The industrial usage of COSMO-RS got further impact by its success in the first Industrial Fluid Property Simulation Challenge (IFPSC) (67, 68), where COSMO-RS outperformed force-field based simulations methods in the prediction of previously unpublished VLE data. In the 5<sup>th</sup> IFPSC two COSMO-RS contributions, one using COSMO-SAC and the other COSMO*therm*, were ranked first and second. Besides the industrial user group, the number of academic chemical

engineering institutes using COSMO-RS for LLE, VLE and SLE applications is continuously growing, reflected by a broad variety of applications (e.g. 69 - 71 ).

Since 2002 COSMO-RS gained large attention in the field of ionic liquids, where it proved to be able to make reasonably accurate predictions for the activity coefficients of solutes in ionic liquids, without any adjustments or re-parameterization. Since all other methods are much harder to apply to ionic liquids, COSMO-RS meanwhile is widely used for the screening of suitable ionic liquids for certain separation and reaction problems (33,72 – 80).

Beyond such more or less standard VLE, SLE, and LLE applications, COSMO-RS has been used for the simulation of more demanding separation and reaction systems, as polymer membranes, activated carbon, micellar systems, hyper-branched polymers, and recently as well for different chromatographic separation process as HPLC and reversed phase HPLC, and SILP (51,81 - 85). Many of these innovative applications have been done by the group of Arlt.

A second important application area and user group has built up in the field of consumer and personal care product research, i.e. in the area of cleaners, shampoos, soaps, perfumes, etc. In this field it is used for the calculation of solubilities and partition behaviour of the various components in the often very complex liquid phases and also for the estimation of the binding of the ingredients to target phases as skin, hair, cotton, and many more (53,86). In this context the  $\sigma$ -moment approach and the COSMOmic extension are often very useful. Since many of the very complex phases are often simulated with mesoscopic simulation techniques as MESODYN or DPD ( 88), quite recently COSMO-RS has started to be used for the estimation of the interaction parameters required for such simulations (89).

Another important application area is pharmaceutical and agrochemical research. COSMO-RS is used in this area for solubility predictions,  $pK_a$  predictions, various physiological partition properties including  $\sigma$ -moment approaches, and for drug similarity studies (20, 57, 58, 90, 91). Nevertheless, the technically most important application in this sector appears to be the solvent screening in the early drug development phase, where it can provide robust predictions of the relative solubility of the drug candidate in a large number of solvents and thus helps to identify the most promising solvents for purification and re-crystallization without time-consuming and often expensive experiments. (92 - 94)

Finally it is worth mentioning the application area of environmental research, which indeed had been the very early starting point for the development of COSMO and COSMO-RS. Reasonable estimates of vapour pressures, partition coefficients, and  $pK_a$  values of most diverse compounds in often very complex environmental phases or at interfaces are of crucial importance for the estimation of the fate of chemical compounds entering the environment in the one or other way. While simple classification and group contribution methods are still being most widely accepted in this politically sensitive area, COSMO-RS is getting increasing attention in this context due to its broader and much more fundamental and more general applicability to a wide range of the environmentally relevant compounds and properties. (95-99) An interesting example is the  $pK_a$  of the industrially and environmentally very important perfluoro-octanoic acid (PFOA), for which Goss and Arp using COSMOtherm predicted a  $pK_a$  value of 0.7, while the experiments and some classification methods seemed to suggest a value in the range of 3 (100). Just recently newer experiments appear to confirm the lower value of the COSMO-RS predictions, leading to a substantially different picture for the environmental transportation mechanism of PFOA.

## 9. Limitations



Despite its very broad applicability, COSMO-RS as any other model obviously has its limitations. The most annoying and unexplainable limitation of COSMO-RS is its inability to correctly represent the interactions of secondary and tertiary amines with hydrogen bonding solutes or solvents.

Another important limitation is given by the accuracy of the calculation of the chemical potentials, which even optimistically should not be expected to be better than  $0.8 \text{ kJ mol}^{-1}$ , without a perspective to become much more accurate in near future. This clearly excludes COSMO-RS from many applications in process development which require much better accuracy. In foreseeable future such high accuracies cannot be provided by any a priori predictive method, but require group contribution or other empirical models specially parameterized on accurate experimental data of similar systems.

A general limitation is the already mentioned restriction to incompressible liquids, which can only be overcome by a combination with EoS or by a hole theory extension. Despite of several successful applications to ionic systems, COSMO-RS is not capable of describing general electrolyte thermodynamics. This is partly due the neglect of long range ion-ion interactions and partly due to the extreme polarization charge densities  $\sigma$  appearing on small, highly charged ions, for which the approximations made in the relative simple misfit and hydrogen bond interactions derived on neutral compounds may no longer be sufficiently accurate. As a general rule it can be said, that ionic systems can be described the better with COSMO-RS, the better the charge is delocalized, i.e. the smaller the surface polarization charge densities are. This explains the success for ionic liquids, which usually have very well delocalized charges. Another limitation arises from the restriction to electrostatic, hydrogen bonding, and dispersive interactions, which excludes the treatment of many important

transition metal complexes, which may interact via orbital interactions or charge transfer with surrounding solvent molecules.

While single strong hydrogen bond interactions can be straightforwardly treated by COSMO-RS without the need for additional association corrections, COSMO-RS cannot by itself describe systems in which a geometrical coordination of strong interactions is of importance, as in carboxylic acid dimerization, complexation to crown ethers or to other chelates, or in drug-receptor binding. In many such cases, as well as in many cases of strong ions and metal complexes, the limitations of COSMO-RS can be overcome by inclusion of one or several of the solvent molecules in a solute-solvent cluster, treating this by QM/COSMO first, and by COSMO-RS in a second step. But such cluster-COSMO-RS applications definitely require quantum chemical and thermodynamic expertise and are definitely not recommended for occasional COSMO-RS.

With respect to properties, COSMO-RS itself is restricted to fluid phase equilibrium properties and cannot describe any dynamic, transport or structural properties of liquid systems. Nevertheless, some of the equilibrium information derived from COSMO-RS may be useful to estimate other properties, as has been shown in the case of viscosity and density predictions. Finally it should be kept in mind that COSMO-RS usually is parameterized mainly on room-temperature data and vapour pressure information. Hence it is most accurate in the temperature range of  $-50^{\circ}\text{C}$  –  $200^{\circ}\text{C}$  and should be used with caution outside this range.

## **Conclusions and outlook**

During the past 15 years the COSMO-RS method has been established as an additional research tool for the investigation of molecular thermodynamics in liquid systems, filling a

gap between the simplifying group-contribution based approaches and the very demanding force-field based molecular simulation methods. While being a bit less accurate than group contribution methods based on experimental data in their core region, the special strengths of COSMO-RS are its robustness and its extrapolation potential into new areas of chemistry and chemical engineering and towards properties not originally considered during the development of COSMO-RS, which result from the theoretically sound combination of quantum chemistry with statistical thermodynamics. Due to its predictive potential COSMO-RS is an ideal tool for the generation and validation of new ideas and for the screening of new alternatives in the early stage of research projects.

Beyond the core application area of bulk fluid phase equilibria, meanwhile a number of extensions towards a broader range of application areas have been developed. One direction consists in  $\sigma$ -moment based QSAR extensions for complex partition properties as adsorption to activated carbon, blood-brain partitioning and other medicinal and biochemical partition parameters. Other examples are the extensions towards inhomogeneous systems as interfaces, micelles, and bio-membranes, COSMOmic, the high throughput module COSMOfrag, and the related extension towards drug similarity screening, COSMOsim.

A special strength of COSMO-RS is its conceptual simplicity and graphicness. Starting from the molecular  $\sigma$ -surfaces resulting from the initial quantum chemical COSMO calculations, the relevant thermodynamic conclusions can be intuitively drawn from the  $\sigma$ -profiles and  $\sigma$ -potentials, ending up with entire phase diagrams. Since all the relevant interactions and thermodynamic phenomena as polar interactions, hydrogen bonding, hydrophobicity, entropy and enthalpy are well covered within this simple concept, it is very well suited as an educational tool for molecular thermodynamics, allowing the students to qualitatively and

quantitatively infer the pathway from molecules to phase diagrams, and to understand, why some molecules like each other, while others do not.

## ACKNOWLEDGMENTS

We gratefully acknowledge the invaluable input of many colleagues and researchers, who contributed to the successful development of COSMO-RS by new ideas, controversial discussions, or innovative applications, but have not been mentioned in this review. A.K. likes to express his gratitude to his former employer Bayer AG, and especially to his former supervisor Peter Hoever, who gave him the scientific degrees of freedom to invent and develop COSMO and COSMO-RS in an industrial research environment.

literature cited,

- [1] Sandler SI., 1999, *Chemical and Engineering Thermodynamics*, New York: Wiley, 3rd Edition
- [2] Hansch C., Leo AJ. 1979. *Substituent Parameters for Correlation Analysis in Chemistry and Biology*, New York: Wiley
- [3] Fredenslund A., Gmehling J., Rasmussen P. 1977. *Vapor Liquid Equilibria Using UNIFAC*, Amsterdam: Elsevier
- [4] van Gunsteren WF., Berendsen HJC. 1990. *Angew. Chem.* 29:992-1023.
- [5] Errington J.R., Panagiotopoulos A.Z. 1999. *J. Phys. Chem. B* 103:6314–22.
- [6] Martin MG., Siepmann JI. 1998. *J. Phys. Chem. B* 102:2569–77.
- [7] Anderson KE., Siepmann JI. 2007. Molecular Simulation Approaches to Solubility. In *Developments and Applications in Solubility*, ed. TM. Letcher. Cambridge:RSC Publishing
- [8] Tomasi J., Mennucci B., Cammi R. 2005. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* 105:2999–3094.
- [9] Klamt A., Schüürmann G. 1993. *J. Chem. Soc. Perkin Trans. 2*:799–805
- [10] Tomasi J., Mennucci B., Cancès E. 1999. *THEOCHEM* 464:211-26
- [11] Chipman DM. 2002. *Theor. Chem. Acc.* 107:80-89
- [12] Klamt A. 1995. *J. Phys. Chem.* 99:2224–35
- [13] Klamt A. 2005, *COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*, Amsterdam: Elsevier
- [14] Klamt A., Jonas V., Bürger T., Lohrenz J. 1998. *J. Phys. Chem.* 102:5074–85.
- [15] This figure was originally published in Reference 13. Copyright Elsevier 2005
- [16] Larsen BL., Rasmussen P. 1986 *Fluid Phase Equilib.* 28:1–11.
- [17] Klamt A. 2003. *Fluid Phase Equilib.* 206:223–35
- [18] Ben Naim A. 1987. *Solvation Thermodynamics*, New York: Plenum Press
- [19] Lin ST., Sandler SI. 2002. *Ind. Eng. Chem. Res.* 41:899–913.
- [20] Klamt A., Eckert F., Hornig M., Beck M., and Bürger T. 2002. *J. Comp. Chem.* 23:275–81
- [21] Gmehling J., Li J., Schiller M. 1993. *Ind. Eng. Chem. Res.* 32:178–93
- [22] Bondi A. 1964, *J. Phys. Chem.* 68:441–51

- [23] Pascual-Ahuir JL., Silla E. 1990. *J. Comp. Chem.* 11:1047-60
- [24] Barone VM., Cossi M. 1998. *J. Phys. Chem. A* 102:1995–2001.
- [25] Eckert F., Klamt A. 1999 – 2009. COSMOtherm, COSMOlogic, Leverkusen, Germany
- [26] Delley B. 1990. *J. Chem. Phys.* 92:508–17
- [27] Delley B. 1992. *J. Chem. Phys.* 94:7245–50.
- [28] DMol<sup>3</sup>, Accelrys, San Diego, CA
- [29] Ahlrichs R., Bär M., Häser M., Horn H., Kölmel C. 1989. *Chem. Phys. Lett.* 162:165–69
- [30] Dewar MJS., Zoebisch EG., Healy EF., Stewart JJP. 1985. *J. Am. Chem. Soc.* 107:3902-09
- [31] Stewart JJP. 1993. A General Molecular Orbital Package, QCPE Program No. 455
- [32] Grensemann H., Gmehling J. 2005. *Ind. Eng. Chem. Res.* 44:1610-24.
- [33] Banerjee T., Singh MK., Khanna A. 2006. *Ind. Eng. Chem. Res.* 45:3207-19
- [34] Pye CC., Ziegler T., van Lenthe E., Louwen JN. 2009 *Can. J. Chem.* 87:790–97
- [35] Klamt A., Mennucci B., Tomasi J., Barone VM., Curutchet C., Orozco M., Luque J. 2009, *Acc. Chem. Res.* 42:489–92
- [36] Marenich AV., Olson RM., Kelly CP., Cramer CJ., Truhlar DG. 2007 *J. Chem. Theory Comput.* 3:2011–33
- [37] Diedenhofen M., Klamt A. 2008. . COSMOconf, COSMOlogic, Leverkusen, Germany
- [38] Buggert M., Cadena C., Mokrushina L., Smirnova I., Maginn EJ., Arlt W. 2009. *Chem. Eng. & Techn.* 32:977-86
- [39] Eckert F., Klamt A. 2000 – 2009. COSMObase, COSMOlogic, Leverkusen, Germany
- [40] Mullins E., Oldland R., Liu YA., Wang S., Sandler S.I., Chen CC. Zwolak M., Seavey KC. 2006. *Ind. Eng. Chem. Res.* 45:4389-4415
- [41] C. Panayiotou, *Ind. Eng. Chem. Res.* 2003, 42, 1495–1507.
- [42] Spuhl O., Arlt W., Del Rio Hernandez A. 2003. *Chem. Ing. Tech.* 75:58-62
- [43] Constantinescu D., Klamt A., Geană D. 2005. *Fluid Phase Equil.* 231:231-38
- [44] Leonhard K., Veverka J., Lucas K. 2009. *Fluid Phase Equil.* 275:105 – 15.
- [45] Shimoyama Y., Iwai, Y., Takada S., Arai Y., Tsuji T., Hiaki T. 2006. *Fluid Phase Equil.* 243:183-92

- [46] Hsieh CM., Lin ST. 2008. *AIChE Journal*, 54:2174-81
- [47] Shimoyama Y., Iwai, Y. 2009. *J. Supercrit. Fluids* 50:210-17
- [48] Klamt A., Huniar U., Spycher S., Keldenich J. 2008. *J. Phys. Chem. B* 112:12148–57
- [49] Zissimos AM., Abraham MH., Klamt A., Eckert F., Wood J. 2002. *J. Chem. Inf. Comput. Sci.* 42:1320–31
- [50] Wichmann K., Diedenhofen M., Klamt A. 2007. *J. Chem. Inf. Model.* 47:228–33
- [51] Mehler C., Klamt A., Peukert W. 2002. *AIChE Journal* 48:1093-99
- [52] Klamt A., Eckert F., Hornig M. 2001. *J. Comp.-Aid. Mol. Design* 15:355-65
- [53] Sonnenberg S., Finke A., Klamt A., Lohrenz J., Bürger T., Matthiesen S. 2001. US Patent 6741954
- [54] Delley B. 2006. *Molecular Simulation* 32:117-23
- [55] Klamt A. 2008. *Prediction of the Free Energies of Crystal Faces In Pure and Mixed Solvents by COSMO-RS*. Presented at Annual Meeting AIChE, Philadelphia
- [56] Hornig M., Klamt A. 2005. *J. Chem. Inf. Model* 45:1169–77
- [57] Thormann M., Klamt A., Hornig M., Almstetter M. 2006. *J. Chem. Inf. Model* 46:1040–53
- [58] Klamt A., Eckert F., Diedenhofen M., Beck ME. 2003. *J. Phys. Chem. A* 107:9380–86
- [59] Klamt A., Eckert F. 2006. *J. Comp. Chem.* 27:11-19
- [60] Eckert F., Leito I., Kaljurand I., Kütt A., Klamt A., Diedenhofen M. 2009 *J. Comp. Chem.* 30:799-810
- [61] Clausen I. 2000. *Experimentelle und theoretische Analyse der Anwendbarkeit von COSMO-RS für Verteilungskoeffizienten* PhD thesis. Techn. Univ. Berlin, Berlin
- [62] Clausen I., Arlt W. 2002. *Chem. Eng. & Technol.* 25:254-58
- [63] Schäfer A. 2003. *Chem. Ing. Tech.* 75:1000-00
- [64] Putnam R., Taylor R., Klamt A., Eckert F., Schiller M. 2003. *Ind. Eng. Chem. Res.* 42:3635–41
- [65] Ruffert G., Olf G. 2006. *Chemie Technik* 33:86-88

- [66] Franke R., Krissmann J., Janowsky R. 2002. *Chem. Ing. Tech.* 74: 85-89
- [67] Case F., Chaka A., Friend DG., Frurip D., Golab J., Johnson R., Moore J., Mountain RD., Olson J., Schiller M., Storer J. 2004. *Fluid Phase Equil.* 217:1-10
- [68] Klamt A., Eckert F. 2004. *Fluid Phase Equil.* 217:53-57
- [69] Milocco O., Fermeglia M., Pricl S. 2002. *Fluid Phase Equil.* 199:15-21
- [70] Fermeglia M., Braiuca P., Gardossi L., Pricl S., Halling PJ. 2006. *Biotech. Prog.* 22:1146-52
- [71] Peters M., Greiner L., Leonhard K. 2008. *AIChE Journal* 54: 2729-34
- [72] Marsh KN., Deev A., Wu ACT., Tran E., Klamt A. 2002. *Kor. J. Chem. Eng.* 19:357-62
- [73] Diedenhofen M., Eckert F., Klamt A. 2003. *J. Chem. Eng. Data* 48:475-79
- [74] Jork C., Kristen C., Pieraccini D., Stark A., Arlt W. 2005. *J. Chem. Therm.* 37:537-58
- [75] Kato R., Gmehling J. 2005. *J. Chem. Therm.* 37:603-19
- [76] Guo Z., Chen B., Murillo RL., Tan T., Xu X. 2006. *Org. Biomol. Chem.* 4:2772 -76
- [77] Lei ZG., Arlt W., Wasserscheid P. 2006. *Fluid Phase Equil.* 241:290-99
- [78] Zhao H. 2006. *Chem. Eng. Comm.* 193:1660-77
- [79] Palomar J., Ferro VR., Torrecilla JS., Rodriguez F. 2007. *Ind. Eng. Chem. Res.* 46:6041-48
- [80] Han TYJ., Pagoria PF., Gash AE., Maiti A., Orme CA., Mitchell AR., Fried LE. 2009. *New J. Chem.* 33:50 – 56
- [81] Diedenhofen M., Klamt A., Malsch G., Frahn J., Friedl A. 2002 *COSMO-RS Property Predictions for the Directed Development of Polymer Membranes* Presented at 17th IUPAC Conference on Chemical Thermodynamics, Rostock
- [82] Mokrushina L., Buggert M., Smirnova I. , Arlt W., Schomaker R. 2007. *Ind. Eng. Chem. Res.* 46:6501-09
- [83] Arlt W., Spuhl O., Klamt A. 2004. *Chem. Eng. & Proc.* 43:221-38



- [84] Reithinger M., Arlt W. 2009. *Prediction of the Partition Coefficient in Liquid-Solid Chromatography Using COSMO-RS* Presented at 8th World Congress of Chemical Engineering (WCCE8); Montreal, Canada
- [85] Buchele A., Arlt W., Haumann M., Wasserscheid P. 2009 *Modeling of the Supported Ionic Liquid Phase (SILP) Catalysis* Presented at 8th World Congress of Chemical Engineering (WCCE8); Montreal, Canada
- [86] Verstraete P., Murch B. 2009. *Use of COSMOtherm in Modeling Consumer Product Compositions* Presented at COSMO-RS Symposium 2009, Wermelskirchen, Germany
- [87] Laidig DW., Enke DM. 2009 *Modeling systems for consumer goods* US Patent Application 20090094006
- [88] Glotzer SC., Paul W. 2002. *Ann. Rev. Materials Res.* 32: 401-36
- [89] Fermeglia M., Priol S. 2009. *Computers & Chem. Eng.* 33:1701-10
- [90] Thormann M. 2009. *COSMO-RS & COSMOsim - Valuable Tools for Drug Discovery* Presented at COSMO-RS Symposium 2009, Wermelskirchen, Germany
- [91] Beck ME. 2009. *Applications of COSMO-RS in Crop Science Research* Presented at COSMO-RS Symposium 2009, Wermelskirchen, Germany
- [92] Abramov Y., Mustakis J., am Ende D. 2009 *COSMO-RS applications in Pharmaceutical science* Presented at COSMO-RS Symposium 2009, Wermelskirchen, Germany
- [93] Geertman R. 2009. *COSMOtherm: The ideal solvent toolbox?* Presented at COSMO-RS Symposium 2009, Wermelskirchen, Germany
- [94] Deij M. 2009. *Using COSMOtherm in high-throughput polymorph screening of pharmaceuticals* Presented at COSMO-RS Symposium 2009, Wermelskirchen, Germany
- [95] Klamt A., Eckert F., Diedenhofen M. 2002. *Environ. Tox. Chem.* 21: 2562–66
- [96] Niederer C., Goss KU. 2007. *Environ. Sci. Technol.* 41:3646–52
- [97] Endo S., Xua W., Goss KU., Schmidt TC. 2008. *Chemosphere* 73:532-38
- [98] Oleszek-Kudlak S., Grabda M., Shibata E., Eckert F., Nakamura T. 2005.

*Environ. Tox. Chem.* 24:1368-75

[99] Goss KU., Arp HPH. 2009 *Environ. Sci. Technol.* 43:1923–29

[100] Goss KU., Arp HPH. 2009 *Environ. Sci. Technol.* 43:5150–51

[101] MacDonald SM., Opallo M., Klamt A., Eckert F., Marken F. 2008. *Phys. Chem. Chem. Phys.* 10:3925-33

Article components:

**Keywords:**

chemical potentials, activity coefficients, partition coefficients, solubility, mixture  
thermodynamics, ionic liquids

**Abstract**

The COSMO-RS method has been established as a novel access to the prediction of thermophysical data of liquid systems and has become a frequently used alternative to force-field based molecular simulation methods on the one side and group-contribution methods on the other side. By its unique combination of a quantum chemical treatment of solutes and solvents with an efficient statistical thermodynamics of the molecular surface interactions it enables the efficient calculation of many properties which can hardly be predicted by other methods. This review presents a short delineation of the theory, its application potential and limitations, and of its most important application areas.

**Terms/Definitions list:**

if the editor has suggestions which terms should be defined, please let us know.

**Acronym list:**

NRTL	Non-Random-Two-Liquids model
UNIQUAC	UNIversal QUAsi-Chemical model
CLOGP	Calculated LOGarithmic octanol-water Partition coefficients model
UNIFAC	Universal Quasichemical Functional Group Activity Coefficients model
MD	Molecular Dynamics sampling technique in molecular simulations
MC	Monte-Carlo sampling technique in molecular simulations
QC	Quantum Chemistry, i.e. quantum mechanics applied to chemistry
CSM	Continuum Solvation Models
COSMO	COnductor-like Screening MOdel

IEFPCM	Integral Equation Formalism Polarizable Continuum Model
SS(V)PE	Surface and Simulation of Volume Polarization for Electrostatics model
COSMO-RS	COnductor-like Screening MOdel for Realistic Solvation
DFT	(electron-)density functional theory
MP2	Moeller-Plesset 2 <sup>nd</sup> order perturbation theory
$\sigma, \sigma'$	conductor polarization charge density on the surface of a molecules
OCE	Outlying Charge Error
$a_{\text{cont}}$	contact surface segment of two molecules
$E_{\text{int}}(\sigma, \sigma')$	interaction energy of surfaces segments with polarization charges $\sigma$ and $\sigma'$
$e_{\text{int}}(\sigma, \sigma')$	surface specific interaction energy of polarization charges $\sigma$ and $\sigma'$
$p^X(\sigma)$	$\sigma$ -profile of a molecules X
$p_s(\sigma)$	$\sigma$ -profile of a solvent or mixture S
$\mu_s(\sigma)$	$\sigma$ -potential of a solvent or mixture S
$\Delta G_{\text{fus}}^X$	Gibbs free energy of fusion, i.e. the free energy difference of a compound between its crystalline and liquid state
VLE	vapour-liquid equilibrium
LLE	liquid-liquid equilibrium
SLE	solid-liquid equilibrium
QSPR	quantitative structure property relationship
RMSE	Root Mean Squared Deviation
MUE	Mean Unsigned Error
MESODYN	MESOscopic DYNamics simulation method
DPD	Dissipative Particle Dynamcis, a mesoscopic simulation method
PFOA	PerFluoro-Octanoic Acid

**Summary Points list:**

- Quantum chemical calculations of molecules in a virtual conductor (COSMO) yield valuable reference information for molecules in the liquid phase.
- COSMO polarization charge densities can be used to quantify molecular surface interactions.
- COSMO-RS statistical thermodynamics of surface pairs yields converts the surface interactions into chemical potentials, and these lead to all equilibrium related liquid phase thermophysical properties.
- COSMO-RS is widely used for in process design and development, for solvent and solubility screening, including ionic liquids.
- COSMO-RS can also be used for the exploration of more advanced topics as  $pK_a$ -prediction, partition behaviour at liquid interfaces, in chromatographic systems, in micelles and in bio-membranes, partitioning and adsorption involving complex pseudo-phases, and even for similarity evaluation in drug design.
- COSMO-RS provides a rather graphic bridge from molecules to liquid phase thermodynamic properties and phase behaviour, which can be very well used for teaching molecular thermodynamics.

**Future issues list:**

- For an overall improvement of the accuracy of COSMO-RS a combination with a more accurate, but still computationally affordable, quantum chemical method than DFT is required.
- More detailed and accurate  $\sigma$ -based expressions for the hydrogen bond interactions need to be developed.

- A more sophisticated description of ion interactions, including long-range ion-ion interactions and better expressions for strong surface interactions, is required for a more general applicability of COSMO-RS to ionic and electrolyte systems.

### **Annotated References:**

13 This book gives the most detailed description of the COSMO-RS method currently available.

### **Side Bar:**

Prediction of the free energy of transfer of ions

A demanding challenge for COSMO-RS was presented by Frank Marken in 2007. He wanted to compare the free energies of transfer of a very diverse set of ions measured by voltammetry with the predictions by COSMO-RS. The set of ions consisted of 18 anions, ranging from simple ions as  $\text{Cl}^-$  and  $\text{BF}_4^-$  to highly flexible organic ions as singly, doubly, triply charged citrate. The ion transfer was measured from water to the organic solvent 4-(3-phenylpropyl)pyridine (PPP). The initial comparison showed a very good correlation between the experimental and calculated transfer free energies, but a closer look disclosed a systematic overestimation of the transfer free energies by a factor of 2. As a possible explanation of this overestimation we suggested the water content of the organic phase, which the experimentalists were not aware about. COSMO-RS predicted a considerable water content of 45 mol% for water saturated PPP. An experimental determination even yielded a slightly larger water content of 55 mol%. Including the water content of PPP into the COSMO-RS calculations of the free energies of transfer of the ions now yielded a good regression, now with the expected correct slope of the experimental vs. calculated energies

(101). Hence COSMO-RS elucidated the crucial role of small water contents for the solvation of ions.